

2984 words

## Haig's 'strange inversion of reasoning'

Daniel C. Dennett

Tufts University

### Abstract

David Haig (this issue) propounds and illustrates the unity of a radically revised set of definitions of the family of terms at the heart of philosophy of cognitive science and mind: *information, meaning, interpretation, text, choice, possibility, cause*. This biological re-grounding of much-debated concepts yields a bounty of insights into the nature of meaning and life.

### 1. The View From The Top Down

---

Acknowledgments: I am indebted to David Haig, Enoch Lambert, Rosa Cao, Nicholas Shea and Jura Pintar for enlightening discussions of these ideas.

Address for correspondence: Center for Cognitive Studies, Tufts University,  
Medford, MA 02155

email: [daniel.dennett@tufts.edu](mailto:daniel.dennett@tufts.edu)

One of the most vehement 19<sup>th</sup> century critics of *Origin of Species*, William MacKenzie Beverley (1868), summed up Darwin's contribution brilliantly:

In the theory with which we have to deal, Absolute Ignorance is the artificer; so that we may enunciate as the fundamental principle of the whole system, that, IN ORDER TO MAKE A PERFECT AND BEAUTIFUL MACHINE, IT IS NOT REQUISITE TO KNOW HOW TO MAKE IT. This proposition will be found, on careful examination, to express, in condensed form, the essential purport of the Theory, and to express in a few words all Mr. Darwin's meaning; who, by a *strange inversion of reasoning* [my emphasis], seems to think Absolute Ignorance fully qualified to take the place of Absolute Wisdom in all the achievements of creative skill.

I have adapted Beverley's all-caps howl to describe what I call Turing's strange inversion of reasoning (Dennett, 2013a):

IN ORDER TO BE A PERFECT AND BEAUTIFUL COMPUTING MACHINE IT IS NOT REQUISITE TO KNOW WHAT ARITHMETIC IS.

Both Darwin and Turing turned our everyday mind-first perspective upside down, shocking many who just could not—at first—countenance their joint message: there can be *competence without comprehension* (Dennett, 2017). I assume that the readers of *Mind and Language* have already made their peace with these newish

perspectives (though some thoughtful people still ardently resist them); I want to introduce them to yet another unsettling inversion, which may be harder to take—at first.

The title of this journal, *Mind and Language*, seems straightforward enough, but it exemplifies a tacit bias that has gone largely unchallenged for more than fifty years. We might express it thus:

When it comes to Meaning, the best starting place is (obviously) Language, and the way Minds use language.

Whether meaning resides first and foremost in Minds (*e.g.*, the ‘original intentionality’ of Searle (1980)) or in linguistic structures (*e.g.*, a natural language or the Language of Thought championed by Fodor, 1975), investigations into semantics—and information, interpretation, belief and knowledge—should begin by exploiting what we have already learned about human minds and human language. Why? Because when we examine the ways meanings are composed in our own minds and in our languages we get to see some of the inner workings of a machine (an organ, a system, a phenomenon, . . . .) that captures, modulates and transmits meaning. We get to catalogue the meaningful parts out of which larger meanings are composed in *these* systems (which might just be a single system with different manifestations), and this must surely (*ding!*)<sup>1</sup> give us insight into all other systems in which meanings get extracted, modulated, composed and transmitted.

---

<sup>1</sup> The ‘surely’ alarm was introduced in Dennett (1994) and elaborated in Dennett 2013, pp53-4. Readers are encouraged to inculcate the habit of hearing a bell ring in

This is the consensual justification for setting up our base camp in Mind and Language, a secure and literal arena of meanings from which to venture into more exotic and perhaps merely metaphorically meaning-ful phenomena, such as the behavior of other animals, or plants, or robots, or life itself. *Mind and Language* is a journal of philosophy and linguistics (primarily), and certainly not a journal of theoretical biology. If there is meaning in DNA, for instance, then DNA must be rather like a written language, with readers and writers, a productive (syntactical) system with something like terms that have something like semantics, but of course the genetic system is not *just* like a (human) language; it's a *sorta* language,<sup>2</sup> a language thanks to poetic license or metaphor.

This familiar—indeed normal—assumption shares with the pre-Darwinian and pre-Turingian worldview the property of *taking comprehension for granted*: if you are talking about meaning and communication, you are talking about comprehending agents, human adults, paradigmatically. This may seem to be an obvious starting point, but it is an unheralded partisan in the embattled gulf often proclaimed (and often deplored) between the *Naturwissenschaften* and the *Geisteswissenschaften*, or between C. P. Snow's Two Cultures. Among its ideological buttresses are Brentano's thesis of the Irreducibility of the Intentional, which obliged Quine (1960) to take sides, at least temporarily, against the Intentional

---

their heads whenever they encounter the word 'surely'—since it often marks the weakest point in the argument.

<sup>2</sup> I deliberately use the 'sorta operator' I defined and defended in Dennett (2013) to implicate myself in this bias, which I am now attempting to expose. The sorta operator is a valuable tool in any naturalist's kit, but like all powerful tools, it can be used in ways that are not constructive.

realm, and Chomsky's (1959) fabled demolition of Skinner's *Verbal Behavior* (1957). Behaviorism, both Quine's and Skinner's if not Ryle's or Wittgenstein's, has been banished (*Hurrah!*) to make the world safe for Cognitive Science, the properly scientific study of Mind and Language. What this now canonical presumption ignores is a different sort of behaviorism, more fundamental and less tinged with over-ambitious ideology: the behaviorism of science. Meteorology is behavioristic in this sense, and so is chemistry, and physics and geology and astronomy. When you achieve a theory that explains all meteorological *behavior*, you get to declare victory; you've finished the task, because that's all there is.

What about biology? Here there is still an ominous hint of a door left open to dualism or its near-twin vitalism. The standard, if defeasible, assumption in biology these days is that in the end there are just lots of macromolecules and their constituent atoms, and the forces that govern their *behavior*, even if it is extremely useful to find higher levels of abstraction from which to search for patterns: the theory of the cell, for instance, and molecular genetics, and various grain-levels of neuroscience, evolutionary biology and ecology. There is no *élan vital* or ectoplasm to account for, so when the 'behavioral and brain sciences' accomplish their tasks, there will be no residual hard problems to solve. Or will there? The perspective adopted by most in cognitive science remains agnostic about this, pending further discoveries about mind and language. In the meantime, armed with our theories and concepts—intension, extension, Boolean operators, propositions, codes, syntax, etc., etc.—we can explore the realm of meaning and information-processing in the brain (or mind) without settling the issue of whether it all somehow 'reduces to' the

behavior of molecules. We have learned, haven't we, that we can speak about information—thanks to Shannon's mathematical theory of information, Marr's (1982) 'computational level,' Newell's (1982) 'knowledge level'—in advance of any firm anchoring in the behavior of neurons and such. I say 'we' because I have not, until now, appreciated how I myself have participated and even contributed to this comfortable *modus operandi*.

## **2. Starting With Life, Not Language**

What would it look like to abandon this normal perspective on mind and language and meaning, and try to reframe all the issues at one biological level or another, building *up* to the niceties of yearning and doubt, prosody and implicature from the ground floor of meaning, in the origin of life itself? I commend to the readers of *Mind and Language* a pioneering essay by evolutionary biologist (and amateur philosopher, in the best sense of both terms) David Haig, 'Making Sense: information interpreted as meaning,' (this volume), which boldly outlines just such a project. It will upset any reader who is sure that the 'proper' meanings of the key terms are being flouted or ignored. *Real* interpretation, *real* meaning, is not to be found in the reactions of molecules to other molecules, but in the reactions of sentient intelligences to perceptions and communications. I urge that such reactions be stifled for the time being, until Haig's alternative perspective can be seen in action, a strange but beautiful inversion of reasoning with a bounty of insights. (It reminds

me of the fabled qualms about admitting zero as a number; whatever your misgivings, give it a chance and see if you like the results.)

Haig is not by any means the first to try to ground meaning in biology, turning away from what we might call the linguicentric tradition. Ruth Millikan's pioneering work (1984) has led to a variety of 'teleosemantic' theories (Papineau, 1998, Shea, 2007, and others). Paul Churchland (1995, 2012) on vector coding, Brian Skyrms (2010) on signals, and Terrence Deacon (1997, 2011) are all important contributors to this centrifugal agenda, and some of my own work (e.g., 2001) has concentrated on exposing and resisting the siren song of the 'propositional attitude task force.' Lean, 2014, develops congenial points, but Haig is the first, to my knowledge, to simply defy tradition and propose alternative readings of the standard terms, *information, choice, meaning, intention, interpretation, text*, and thereby everything that we build from them. This defiance can perhaps be most directly motivated by noting that none of these terms, however ubiquitous in theories and models of mind and language, have ever settled down with consensual definitions. Every one of them is an attractor of controversy, so why not start fresh and see what happens?

The governing insight is that we can better understand what 'real' choices are, what 'real' interpretation is, what 'real' meaning is if we consider first the primordial phenomena on which and out of which these must be built: the simplest possible differences that make a difference, the most rudimentary 'choices'

imaginable. This enables Haig to undo the divorce between ‘Shannon information’ and ‘semantic information’ by seeing how to unite the striking features of these warring perspectives into a single account of how information is used in life, including our own highly articulated intellectual lives.<sup>3</sup>

Information is what could have been otherwise before observation. Meaning is what would have been otherwise had the observation been different. (Haig, ms. p24)

A device that strikes a match unless it ‘observes’ (detects) the presence of hydrogen ‘uses’ information to govern its ‘choice’ to strike the match, and is ‘undecided’ until it gets that information and ‘interprets’ that information by striking or not striking the match. Such a simple device could be invented by a (not very) intelligent designer, or it could evolve. Establishing just such a (type of) causal link is the basic or atomic design step out of which all other design-making must be constructed. Where before there was no regular causal link between hydrogen-presence and not-striking-a-match, now there is, and it can be a useful causal regularity worth keeping. Dretske’s (1981, 1994) attention was rightly drawn to such second-order causes of causal regularities, but he was stymied by the task of saying when there was *enough* design, or specificity, or complexity to warrant assigning meaning to the activity involved (for instance, in his attempt to distinguish

---

<sup>3</sup> Dennett (2017) was completed before I had digested Haig’s ideas, and Chapter 6 of that book, ‘What is Information?’, now stands in need of revisions just weeks after appearing in print. This essay is a first installment of that editorial process.



‘epistemic seeing’ from ‘non-epistemic seeing’). Haig doesn’t bother: it may become more and more *convenient* to speak the shorthand of meaning-assignments as systems become more complicated in their causal intermeshing, but there is no point at which, as Fodor (1987, p104) once memorably put it, ‘a whistle blows’ heralding the fixation of Meaning.

Instead of treating the bacterium’s response to the presence of a molecule as a fringe or degenerate or merely metaphorical case of meaning, lacking most of the ‘essential’ components of what we usually mean by meaning, consider treating it as the most fundamental example of meaning: the simplest case. Then, treating *intention* to cover the rationales of evolved designs as literally as the rationales of human designers, we can say, literally, that ‘interpreters are intentional mechanisms that have evolved or been designed to use information in choice,’ (ms p3) and define ‘a *text* as an interpretation intended to inform subsequent choice.’ (ms p8) and go on to say that the mutation that causes hemophilia in males does so via a process in which histidine and hemophilia

‘are the intended meanings of interpreters [ribosomes] that have evolved to represent whatever text they are presented. But, histidine and hemophilia are unintended from the perspective of the organism: none of the fetus’s male ancestors possessed the mutant protein or suffered from hemophilia.’ (ms p13)

The same strategy of analysis yields rich and unified accounts of the tangled history of the Zimmerman telegram, the sincere advice of fortune-tellers, and the

underlying rationale of the injury-feigning bird, among many other phenomena. (Examples are the key; you don't really appreciate the power of definitions until you see the, work in multiple cases.) Then, with these new foundations in place Haig can go on to define a *percept* as a *private text*, 'an interpretation of sensory input intended to inform subsequent interpretation.' (ms p10), and then reflexively characterize his project in its own terms:

My written text attempts to rearrange the associations of 'meaning' and 'information' in your private texts to change how you interpret and use these words. It is an invitation to join a language-game in which these new definitions are the rules of play. (ms p22)

Instead of attempting to summarize Haig's elegant and succinct proposals, I will simply note a few of the philosophical topics on which they shed surprising light:

1. the debate over causation as mechanism versus causation as difference-making,
2. the reason why an epistemic—not metaphysical—concept of possibility does all the heavy lifting in the context of the puzzles about free will,

3. the relation between the butterfly effect and attractor states, and the implications of this for cognition: 'Bodily attractors at all levels—from the molecular, to the cellular, to the individual—buffer organismal fates from the unforeseen. This cancellation of irrelevant differences allows "focus" on what is relevant.'

No doubt my high opinion of Haig's essay is due in part to the fact that it clarifies, extends and corrects points I have long attempted—unsuccessfully—to convey to my colleagues. I have often tried to show that, whether the topic is what the frog's eye tells the frog's brain, or what Hamlet told Ophelia, or what the best translation of a sentence from French to English must preserve, our inability to 'express' the 'content' in some canonic medium of communication or information-registration is not a bug but a feature. Haig shows why and how we may contrive all manner of intermediate levels of expression or interpretation but need not hunt for a dividing line that distinguishes comprehension from mere reaction. Instead of saying, as some in the linguicentric tradition might want to put it, that it's coding and syntax 'all the way down,' we can now appreciate that it's interpretation all the way up. Many of the difficult problems with which theorists of mind and language wrestle remain unresolved, but there is reason to believe that Haig's strange inversion will open up unconsidered prospects that have great promise.

## References

- Beverley, W. M. 1868: *The Darwinian Theory of the Transmutation of Species Examined*. London: Nisbet.
- Chomsky, Noam. 1959: A Review of B. F. Skinner's Verbal Behavior. *Language*, 35, No. 1, 26-58.
- Churchland, Paul. 1995: *The Engine of Reason, the Seat of the Soul*. Cambridge, MA: MIT Press.
- 2012: *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals*. Cambridge, MA: MIT Press.
- Deacon, Terrence William. 1997: *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: W. W. Norton.
- . 2011: *Incomplete Nature: How Mind Emerged from Matter*. New York: W. W. Norton.
- Dennett, Daniel C., 1994: Get Real. *Philosophical Topics*, 22, (1), 505–568.
- 2001: Things about Things. In J. Branquinho, (ed.), *The Foundations of Cognitive Science*. Oxford: Clarendon Press.
- 2013a: Turing's 'Strange Inversion of Reasoning. In S. B. Cooper and J. van Leeuwen (eds.), *Alan Turing: His Work and Impact*. Elsevier.
- 2013b: *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton.

-----2017: *From Bacteria to Bach and Back: the Evolution of Minds*, New York: Norton.

Dretske, Fred, 1981: *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

-----1994: Differences that make no difference. *Philosophical Topics* 22 (1), 41-57.

Fodor, Jerry, 1975: *The Language of Thought*, New York: Crowell.

-----1987: *Psychosemantics : The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

David Haig, (this volume): Making Sense: information interpreted as meaning. (This volume).

Lean, Oliver M. 2014: Getting the Most out of Shannon Information. *Biology & Philosophy*, 29, 395-413.

Marr, David, 1982: *Vision*. Cambridge, MA: MIT Press.

Millikan, Ruth Garrett. 1984: *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, Mass.: MIT Press.

Newell, Allan. 1982: The Knowledge Level. *Artificial Intelligence*, 18, 81-132.

Papineau, David. 1998: Teleosemantics and indeterminacy. *Australasian Journal of Philosophy*, 76 (1), 1-14.

Quine, W. V. O. 1960: *Word and Object*. Cambridge, MA: MIT Press.

Searle, John. R. 1980: Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3 (3), 417-457.

Shea, Nicholas, 2007: Consumers Need Information: supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research* 75, (2), 404-435.

Skinner, B. F. 1957: *Verbal Behavior*. Acton, MA: Copley Publishing Group.

Skyrms, Brian. 2010: *Signals: Evolution, Learning, and Information*, Oxford Univ. Press.

Center for Cognitive Studies, Tufts University

**Making sense:**  
**information interpreted as meaning**

**David Haig**

Department of Organismic and Evolutionary Biology,

Harvard University, 26 Oxford Street,

Cambridge MA 02138.

dhaig@oeb.harvard.edu

**Abstract:** An interpreter is a mechanism that uses information in choice. The capabilities of the interpreter couple an entropy of inputs (uncertainty) to an entropy of outputs (indecision). The first entropy is dispelled by observation (input of information). The second entropy is dispelled by choice of action (output of decision). I propose that an interpreter's response to inputs (information) be considered the *meaning* of the information *for* the interpreter. In this conceptual framework, the designed or evolved mechanisms of interpreters provide the much-debated link between Shannon information and semantics.

*In principio erat finis.*

A mechanical device strikes a match and a candle is lit or there is an explosion. The striking of the match (+M) and the presence of oxygen (+O) are the same in both scenarios. The difference that makes the difference as to whether or not there is an explosion ( $\pm E$ ) is the presence versus absence of hydrogen ( $\pm H$ ). A more sophisticated device strikes a match contingent on input from a hydrogen sensor. If the sensor reports no hydrogen ( $-H$ ), the match is struck (+M). If the sensor reports hydrogen (+H), the match is not struck ( $-M$ ). The first device is an effector of an explosion in the presence of hydrogen but does not 'choose' the explosion because it does not 'use' information. It couples a state of the world ( $\pm H$ ) to an outcome ( $\pm E$ ). The second device 'prefers' darkness to an explosion in the presence of hydrogen. It couples one bit of information about the world ( $\pm H$ ) to one degree of freedom in action ( $\pm M$ ). It 'responds' to +H with  $-M$  and  $-H$  with +M. It is 'undecided' until the observation of what was uncertain (information) is interpreted in definite action (meaning).

The first difference in the phrase 'a difference that makes a difference' is cause, the difference-maker or independent variable, and the second is effect, the difference-made or dependent variable. But whether the second difference is an interpretation of the first depends on the evolved or designed function of an interpreter. The first device does not interpret. Things just happen ( $\pm H$ ,  $\pm E$ ). For the second device, the first difference is information ( $\pm H$ ) and the second is meaning ( $\pm M$ ). For an outside interpreter of the second device,  $\pm H$  and  $\pm M$  contain mutual (or redundant) information: either can be inferred to 'mean the other'. The observer can predict whether or not the match will be struck (meaning) by observing whether or not hydrogen is present (information) or could infer whether or not hydrogen was present (meaning) by observing whether or not the match was struck (information).



On what reasoned grounds can I claim that the second device interprets the presence of hydrogen as a reason not to light a candle but reject the seemingly parallel claim that the first device interprets the presence of hydrogen as a reason to cause an explosion. My argument made an implicit appeal to proper functions (Millikan, 1989). The proper function of the first device is lighting candles. Explosions are unintended consequences. The proper function of the second device is the use of information to decide whether or not to strike a match.

The responses of an interpreter with one degree of freedom of interpretation may seem mechanical and uncomprehending, hardly deserving the ‘meaningful’ label, but any truly sophisticated interpreter will have many degrees of freedom with multiple levels of internal interpretation in which the interpretation of one part of the system is news (and hence information) to other parts of the system. Consider multiple rewirings of a complex device such that each new device responds to the same inputs with different outputs. There is only a shallow sense in which the inputs of these devices are *causes* of the devices’ outputs. An understanding of *how* inputs are interpreted as outputs requires an understanding of a device’s inner workings. An understanding of *why* a device interprets particular inputs as particular outputs requires an understanding of the device’s function and history. We cannot invoke an omniscient homunculus within the system that has an overview of all that the system ‘knows’ (Dennett, 1991). An interpreter cannot ‘know’ what it will choose until it chooses—if it ‘knows’ it has chosen—but an observer can often predict a consistent interpreter’s choices with confidence.

Dewey (1896) recognized that “stimulus and response are not distinctions of existence, but teleological distinctions, that is, distinctions of function, or part played, with reference to reaching or maintaining an end.” One cannot simply draw an arbitrary boundary around part of a complex web of processes, describe all causes crossing the boundary from outside to inside as stimuli, all causes crossing from inside to outside as responses, and all processes within the boundary as interpretation. Interpreters are intentional mechanisms that have evolved or been designed to use information in choice.

## **Teleology of Interpretation**

Meaning and function are intentional terms. *Fighting the good cause* (Haig, 2014) grounded talk of biological function in the unintended teleology of adaptation by natural selection (following Papineau, 1984; Dennett,

1987, 1995; Millikan, 1989; Neander, 1991). Final causes were presented as efficient, even indispensable, summaries of complex concatenations of efficient causes. A token effect cannot precede its token cause. But when one generalizes from causes of tokens to causes of kinds, effect-tokens both precede and succeed cause-tokens in recursive processes. A full causal account of an egg or a chicken contains long series of past chickens and past eggs. An egg is both an effect of a chicken-that-was and a cause of a chicken-to-be.

‘Natural selection’ subsumes all processes by which the environment *selects* a subset *from a set of actual things*. Reproduction replenishes numbers of the diminished subset before the next round of selection. Although nature’s ‘choices’ are unintended, some of her ‘choices’ leave genetic records that allow repetition of that which was ‘chosen’ as intended choices of living beings. Reproductive recursion is rescued from eternal recurrence of the same by the input of new variation via mutation and the shuffling of genetic texts in sexual reproduction. By these processes, recursively selected subsets accumulate information about what worked in the past (Haig, 2014). And what worked was the interpretation of information from the environment in ‘real time’ rather than evolutionary time. As a consequence, the world now abounds with biological interpreters that *select* which differences will make a difference from the myriad potential causes in their environment and *choose* actions *from sets of possible actions* on the basis of observations that could have been different. The mapping of possible inputs to outputs is embodied in the interpreter’s fine structure, with the fit between information and meaning—the efficaciousness of interpretation—derived from past natural selection refined by developmental processes during the interpreter’s life. Information is used, in the sense of Shannon (1948), to refer to the reduction of uncertainty of a receiver provided by a particular input.

To ‘intend’ is to choose for anticipated effects. Two kinds of intentionality can be distinguished. Primary intentionality is the repetition of causes that worked in the past. This is the intentionality of adaptation by natural selection and of conditioned reflexes. Past effects are anticipated to occur again. Secondary intentionality is choice of action after simulation of possible choices and their effects. Simulated effects are anticipated to occur when the action is performed. Secondary intentionality requires an ability to ‘hold in mind’ and evaluate virtual outcomes. Primary intentionality is ‘primary’ in the sense that it evolved before secondary intentionality.

In *Fighting the good cause* (Haig, 2014), information was said to have meaning *for an interpreter* when it was used *to achieve an end*. The present paper simplifies and clarifies by explicitly equating meaning with the interpretation. The action or thing chosen is the meaning *of* the information *for* the interpreter. Information resides in the differences among things that are ‘possible’ until observation of an actual thing. Meaning is the response of the interpreter to the observation and is itself an actual thing that can be observed and used as information by another interpreter. By these definitions, ‘semantic information’ is a contradiction in terms.

## Information and Meaning

An interpreter can be viewed as an input–output device that uses observations to choose actions (Figure 1). *Interpretation* subsumes all internal processes that couple observations (information) to choice of action (meaning). The number of independent things a device could observe can be considered a measure of its *uncertainty* (entropy of observation). The number of independent actions in its repertoire of response can be considered a measure of its *indecision* (entropy of action). Uncertainty is resolved by observation and indecision by choice. Pearl’s (2000) causal models can be considered a class of interpreters in which the *do*-operator, which fixes the value of an input variable, performs the role of observation. Tononi’s (2004) integration of information refers to internal causal processes of interpreters.

An interpreter’s possible inputs are *the things to which it could respond*. Its possible outputs are *the ways in which it could respond*. These capabilities are subjective competences of the interpreter not objective properties of its world. Observations inform, whether the thing observed is ontologically uncertain (undetermined until observation) or epistemically uncertain (determined but ‘unknown’ until observation). Observation of what is epistemically uncertain provides information about prior events. Meanings may be ‘mistaken’ because of malfunction, unanticipated inputs, or because what was once adaptive is now maladaptive. Unintended meanings may be used as information by other interpreters or by the same interpreter in self-reflection.

My purpose in this ‘behaviorist’ account is not to belittle the complexities of interpretation but to argue there is no ghost in the semantics. Information resides in distinguishable things in the interpreter’s world and the meaning of a particular input for the interpreter is simply whatever physical thing (print on paper, sound vibrations, neural states, etc.) is the output of information processing by the interpreter. The

complexities reside in how the inner workings of the interpreter map observations onto actions. There is no non-material domain in which meanings reside outside of physical interpretation. If you protest that this paragraph means more than ink marks on paper or pixels on a screen, then those marks or pixels have been input to a very sophisticated interpreter, your good self, and I thank you for reading.

Consider a paradigmatic small dark something moving across a frog's visual field that causes the frog to stick out its tongue to intercept the thing. If we treat the frog as a black box, photons falling on retinas are information (input); sticking out the tongue is meaning (output). If we were to peek inside the frog's brain, we would find multiple interpretations of interpretations between sensory excitation and motor action. My claim is that each physical state can be considered the meaning of prior information processing and that these states inform subsequent neural states that are themselves new meanings. The frog's visual system interprets incident photons as information about distance, direction of movement, and speed of the speck. These meanings inform subsequent interpretation as motor action. The frog minimizes immediate interpretation of a speck's nature so as not to give a fly time to interpret the frog's intentions (a small dark object in the mouth is worth ten flies that got away). Once the moving something is intercepted, the frog has ample time to interpret whether the thing is food and what kind of food (using oral rather than visual sensors) and to modify its sensory criteria for future protrusions of the tongue.

Any spoken or written claim by a philosopher about what internal states mean to the frog—whether 'fly,' 'food' or 'small moving thing'—is an interpretation of the philosopher and the philosopher's meaning not the frog's meaning. If we were to peek inside the black boxes of philosophers' minds there would undoubtedly be many interpretations of interpretations, meanings of meanings, before keys were struck on keyboards or words spoken or shouted. If a literate frog were to write a memoir of her experience, she might report that she saw the speck as a fly but was mistaken. Her interpretation would be of similar kind to the philosophers' interpretations. An interpreter, even an interpreter of itself, never has direct access to things in themselves but only to information about things.

The claim that meaning is whatever physical thing an interpreter interprets information to mean is a definition, not a claim that all interpretations are equally useful. Some interpretations are 'better' than others because they inform more subsequent interpretations or enable meaningful interpretation of what was previously uninterpretable. Our perceptions have evolved to present useful information about the

world to guide our actions, and our interpretations of words have evolved, from childhood, to make sense of what others are saying. Although information and meaning are defined relative to an interpreter *as subject*, interpreters may aspire or have evolved to interpret information objectively (Lindley 2000).

## **Tools and Texts**

The non-living world is a repository of unintended information useful for living interpreters. Unintended information is also present in interpretations of other interpreters. When an interpretation is reinterpreted, one must distinguish the intentions of the first interpreter (the producer) from those of the second interpreter (the consumer). The evasive movements of a gazelle being pursued by a cheetah are intended to make the gazelle harder to catch. The cheetah observes and interprets these movements with the intention of catching the gazelle. The cheetah's interpretation is unintended by the gazelle.

When a healthy gazelle sees a hunting dog rather than a cheetah, it interprets the situation as an occasion to stot (jump up and down). Hunting dogs preferentially chase gazelles that do not stot or stot more feebly. A vigorous gazelle and a hunting dog both benefit from the hunting dog chasing a more feeble gazelle. The hunting dog's decision to chase a non-stotting gazelle is intended by the stotting gazelle. The evolutionary rationale of stotting is thought to be that stotting 'signals' to the hunting dog that the gazelle has a good chance of outlasting the hunting dog in an extended pursuit and is therefore not worth chasing (FitzGibbon and Fanshawe 1988). But this is an interpretation of behavioral ecologists not, as far as we know, of either gazelles or hunting dogs. Their interpretations are simply stotting and not chasing. (Gazelles do not stot to cheetahs because cheetahs lack endurance but are capable of short bursts of great speed. Gazelles interpret cheetahs as reasons to get far away as quickly and unpredictably as possible.)

Interpretations are *tools* to achieve ends. Some tools are intended to be used as information by subsequent interpreters or by the interpreter itself at some later time. I will use *text* to refer to an interpretation intended to inform subsequent choice. A text is an output of an *author* (producer) intended to be input to a *reader* (consumer). A text anticipates interpretive competence of intended readers. It may be a static object or dynamic performance. By this expanded definition, written documents, works of art, DNA and mRNA, neural activity, and the tape of a Turing machine are all texts. My spoken words are an ephemeral text 'written' in sound intended to be interpreted by listeners. A painting is a persistent text

‘written’ in pigment intended to be interpreted by viewers. Parallel white lines crossing a road are texts intended to be interpreted by pedestrians as places to cross and by motorists as places where pedestrians cross. A peacock’s tail is a text intended to arouse the admiration of peahens. Stotting is a text intended to discourage hunting dogs.

An author’s intended interpretation of a text should be distinguished from the actual interpretation of the reader. A hunting dog who detects unintended evidence of weakness in a gazelle’s performance may chase a stotting gazelle. The author’s intentions are also distinct from how the author intends a text to be interpreted by readers. Some texts are deceptive. The folded wings of a camouflaged moth are intended to be interpreted as ‘not a moth’ by moth-predators but the flash of ‘eye spots’ as the moth unfolds its wings for flight are intended to be interpreted, for a crucial split-second, as ‘eyes of a moth-predator predator’. If these texts are interpreted by moth-predators as intended by the moth, then they have served their purpose and have been interpreted as intended by the author by being misinterpreted by the reader.

A completed nest contains clues about its construction. If birds model their own nest on the nest in which they hatched, then the parental nest informs the construction of the offspring nest. If parents constructed nests in ways that were easily interpreted by offspring and this method of construction was repeated because it enhanced the survival of the parents’ grandoffspring in offsprings’ nests, then the nest is a text of the parents with an intended interpretation by offspring. This example shows that a thing may function both as a simple tool (for holding eggs) and as a text (for instructing offspring). When the Mafia leave the body of an informer in a town square, the murder is both a direct means to an end (removal of an informer) and a text (a warning to potential informers).

The broken-wing display of a ground-nesting bird is a text intended to be interpreted as ‘here is easy prey’ by a predator. The function of the bird’s ‘helpless fluttering’ is to lead the predator away from hidden eggs that are truly easy prey. If the predator recognizes the display as a text, concludes that ‘a nest is near’, and searches for the nest, then the text has failed to be interpreted as intended but the predator has correctly interpreted the situation. If, on the other hand, the bird truly had a broken wing, then its seemingly helpless fluttering would indeed be helpless fluttering, and an attempt to escape, rather than a text with an intended interpretation. The predator would truly interpret the bird’s actions if it concluded

‘here is easy prey.’ If the predator mistakenly interpreted helpless fluttering as ‘a broken-wing display’ and started looking for a nest then the predator would have misinterpreted the situation.

One could imagine a bird with an actual broken wing ‘unconvincingly’ attempting to escape with the intention that the predator interpret its actions as ‘this is a broken wing display therefore the bird is not easy prey but a nest is near’. If the predator started to search for a nest, then the predator would have recognized that the bird’s actions were a text, would have interpreted the text as intended by the bird, but would have misunderstood the bird’s intentions. The text would have achieved the bird’s intention but foiled the intention of the predator who had been ‘deliberately’ misled.

Ultraviolet photons cause damage to unpigmented skin. Some skins respond to ultraviolet exposure by deposition of melanin. For skins that do not respond in this way, ultraviolet photons are not used as information and simply cause unintended damage. For skins that are able to darken, ultraviolet photons are interpreted by the skin as melanin which is employed as a tool to prevent further damage. An observer of tanned skin can infer that the skin has been exposed to the sun and that, depending on context, the skin’s owner works in the fields, rather than at a desk, or has sufficient leisure to spend time at the beach. Because possession of leisure has social value, some people choose to expose their skin to solar lamps so as to be interpreted as possessors of leisure. In this case, tanned skin is a text intended to inform interpretations of intended observers.

An interpretation (a meaning) can be considered to ‘represent’ the interpreted information and to be an ascription of content to the information *by the interpreter*. Representation raises intriguing questions about how internal processes of an interpreter derive meaning from information and why the interpreter has evolved or been designed to interpret information in the way that it does. A text not only ‘represents’ information used by its author in the text’s composition but ‘presents’ information to intended readers. Presentation raises additional questions of how authors anticipate readers’ responses.

Texts do not act directly but indirectly through the interpretations of readers. Texts do nothing, in and of themselves, but one need only consider the *Quran*, *Declaration of Independence*, and *Protocols of the Elders of Zion* to appreciate the differences texts can make in the world.

## Private and Public Texts

The inner workings of complex interpreters involve delegation of tasks among sub-interpreters, each presenting texts for use by other sub-interpreters. These private texts, intended for internal use, all have a physical form. Some are ephemeral. A percept is an interpretation of sensory input intended to inform subsequent interpretation. Others are persistent. A memory is a textual record to be consulted when needed. Consciousness itself is a private text; our mental desktop, a very short-term memory written in we know not what medium, consulted by other sub-interpreters to know where to 'look' for relevant information. The scene that is seen functions as a simplified and constantly updated interpretation of the world that is compared against incoming percepts for detection of relevant differences.

I view a landscape by Claude Monet in which artfully arranged splodges of color on canvas present a scene on the Seine with sailboats and a group of five ducks to one side. Not only do I see the play of light on water but I hear the clink of rigging as the boats rock at anchor. These meanings are my interpretations. My companion sees gulls rather than ducks. As I move in toward the canvas, the things that I see dissolve into ill-formed blobs. The ducks are revealed as no more than five dabs of white paint. The scene that I saw was underdetermined by input from my eyes, but Monet's genius was to suggest, with minimal means, internal sources of information that filled in the picture. He created the illusion of detail that I expected to find when I looked closely. This paragraph does not contain all the information I want to inform your interpretation. By the artful placement of ink on paper I want to suggest paintings and scenes you have seen in order to evoke an "Aha! I see what he means." I particularly want you to see that an author always relies on rich sources and resources in the private texts of his readers for them to make sense of his public text.

(This paragraph was informed by a parallel analysis of a painting by Bellotto in Dennett 2005, p. 65)

My text is the product of multiple drafts of an evolving text. In the process of reading and re-reading, writing and re-writing, I came to understand what I meant and I mean (Haig, 2011). My meaning is the public text that you see, not some nebulous sense in my mind to which the text points. As my aging mind becomes less nimble, I rely more and more on public texts of previous selves as *aides-memoire* of what I wish to mean. What persists in my brain are reworked memories of earlier drafts and regrets about what I wrote poorly. Once a text is published its meanings for readers are untethered from its author's intentions. My meaning becomes your meaning as you read my text.



## Meanings of Genes

Cells contain three very important interpreters that evolved long ago to interpret aperiodic polymers: DNA polymerases complement sense strands of DNA with their antisense strand; RNA polymerases transcribe sense strands of DNA as RNA; and ribosomes translate messenger RNAs (mRNAs) as proteins. For the DNA polymerase, the meaning of a DNA strand is its complement. For the RNA polymerase, the meaning of the DNA strand is an mRNA. For the ribosome, the meaning of the mRNA is a protein. Thus, the same DNA sequence means different things to a DNA polymerase and an RNA polymerase. The information (input) is the same but the meaning (output) is different. DNA and mRNA are texts intended to be interpreted. RNA polymerases also transcribe transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). These are tools to be used in translation of mRNAs, not texts to be further interpreted. Among the texts interpreted by RNA polymerases and ribosomes are instructions for the assembly of RNA polymerases and ribosomes. Interpreter know thyself.

These molecular machines are general-purpose, mindless interpreters of specialized texts. DNA polymerases, in particular, are like monkish scribes in the scriptorium of the Total Library of all possible texts (Borges, 1939) faithfully copying DNA sequences that are both unilluminated and unilluminating. One can posit a Library of Mendel that contains all possible DNA texts of some finite length (Dennett, 1995). Only an infinitesimal subset of texts the length of the human genome can ever have existed. Natural selection has acted in this much more restricted, yet unimaginably vast, library of Darwin that contains all past and present DNA texts. In this library, differences between texts that are still read and texts that are no longer read are a source of information about what has worked and hasn't worked in the past.

What do genes mean? The short answer is whatever physical thing an interpreter interprets them to mean. A DNA strand is an interpretation of its antisense (via one round of DNA replication) and of its sense (via two rounds of DNA replication). But a gene can also be interpreted as an mRNA (via transcription). Thus, a gene means itself for a DNA polymerase but means an mRNA for an RNA polymerase. Complex interpreters can be built up out of simpler interpreters. A DNA strand is interpreted as a protein by the combined interpretative system of an RNA polymerase plus ribosome via two steps of interpretation (transcribed as mRNA then translated as protein). Thus, a gene means a protein for this compound interpreter.

A more contentious claim is that an organism's genes collectively mean the organism. Past organisms have been responsible for the replication and transmission of present genes which are interpreted via complex processes of development to produce present organisms that are responsible for the replication and transmission of future genes. Organisms and their genomes are thus recursively related via primary intentionality. As such, organisms can be considered to interpret their genomes as themselves. This bald statement should not be over-interpreted. Every interpretation of a genomic text as an organism is unique because texts are always interpreted in the *context* of other sources of information. Organisms interpret their genomes in environmental context and not every detail is intended.

The 282<sup>nd</sup> amino acid of factor VIII protein of most humans is arginine, specified by the codon CGC on the 'sense' strand of a *factor VIII* gene. The complementary codon on the 'antisense' strand GCG is the template for transcription of CGC in the mRNA and this triplet is translated as arginine by the ribosome. Cytosine (C) preceding guanine (G) can be chemically modified by attachment of a methyl group to produce 5-methylcytosine (5-meC). Spontaneous deamination of 5-meC converts 5-meC into thymine (T) creating a heritable mutation that changes 'antisense' GCG to GTG which is interpreted as 'sense' CAC by DNA polymerase. RNA polymerase transcribes GTG (DNA) as CAC (mRNA) which is translated as histidine by the ribosome. Factor VIII protein with histidine rather than arginine as the 282<sup>nd</sup> amino acid fails to clot blood causing life-threatening hemophilia in males.

Mutations in DNA that change an amino acid in protein are described as nonsynonymous. DNA polymerases, RNA polymerases, and ribosomes faithfully interpret nonsynonymous mutations without regard for the functionality of the resulting protein. For a DNA polymerase, CAC means GTG (antisense) and CAC (sense); for an RNA polymerase, GTG means CAC; for a ribosome CAC means histidine. These are the intended meanings of interpreters that have evolved to represent whatever text they are presented. But, histidine and hemophilia are unintended from the perspective of the organism: none of the fetus's male ancestors possessed the mutant protein or suffered from hemophilia. A genetic counsellor receives a printout of the DNA sequence of the *factor VIII* gene present in amniotic fluid of a male fetus whose mother is a carrier of a mutant gene copy on one of her X chromosomes. The difference that will make a difference is an A versus G in the middle position of the 282<sup>nd</sup> codon of the *factor VIII* gene. Before viewing the

printout, the counsellor is uncertain about what the data will show. If she reads A then the child will suffer from hemophilia but if she reads G the child will be unaffected.

Mutations in DNA that do not change the amino acid translated by ribosomes are described as synonymous. An evolutionary biologist might use the ratio of non-synonymous to synonymous differences between two sequences to infer whether the sequences have been subject to natural selection. In this case, meaningless differences for an intended interpreter (ribosome) are meaningful differences for an unintended interpreter (biologist).

## **Telegraphing One's Intentions**

“An engineering communication theory is just like a very proper and discreet girl at the telegraph office accepting your telegram. She pays no attention to the meaning, whether it be sad or joyous or embarrassing. But she must be prepared to deal intelligently with all messages that come to her desk.” (Weaver 1949)

The Zimmermann telegram of 1917 conveyed an offer of a military alliance between the German Empire and the Republic of Mexico if the United States of America entered the Great War. In the event of such an alliance, Germany promised to provide financial support to Mexico and recognize Mexican reconquest of Texas, New Mexico, and Arizona.

Great Britain cut all German telegraphic cables to the New World on 5 August 1914. Therefore the German message to Mexico had to be conveyed by an indirect route. A plaintext of the message in German was composed in Berlin by the secretary of state for foreign affairs (Arthur Zimmermann) and then encrypted by his staff using code 7500. The encoded message was passed to the American Embassy in Berlin for transmission by U.S. diplomatic channels to the German Embassy in Washington. The message was sent by cable from Berlin to Copenhagen, from Copenhagen to London, and then by transatlantic cable from London to the State Department in Washington. The encrypted message in code 7500 was passed to the German Embassy in Washington where it was decrypted into German and then re-encrypted in code 13040. Re-encryption was necessary because the German Embassy in Mexico lacked a code book for the more secure code 7500. The message in code 13040 was sent by telegram from the Western Union office in Washington to the Western Union office in Mexico City where it was printed as a typescript and

delivered to the German Embassy in Mexico on January 1, 1917. The typescript was decrypted into German by the ambassador's secretary, before the offer was communicated by the ambassador to the Mexican President in Spanish.

Unbeknownst to the governments of the United States and Germany, the telegraphic signals that included the message in code 7500 had been intercepted by British intelligence in London and the message partially decrypted. The British then requested their agents in Mexico obtain a copy of the telegram in code 13040. The major reason for this step was to hide from the United States government that British intelligence was reading their diplomatic cables. British intelligence decrypted the telegram (it contained little unexpected for them) and then the British government provided the telegram to the American government in its coded form and as versions decrypted into German and translated into English. The message was then released by the pro-war party in the United States to the American press. The popular outcry hastened the entry of the United States into the war on the side of Great Britain and its allies. The British government had decided that the benefit of revealing the telegram to the United States would outweigh the unavoidable information provided to Germany that Britain could read code 13040.

The 'Zimmerman telegram' passed through many texts in many media, with interpretation occurring at each conversion from one medium to another. Sometimes 'the message' was in German, sometimes in code 7500, sometimes in code 13040, sometimes in Morse, sometimes in Spanish, and sometimes in English. For the various clerks in telegraph offices in Berlin, Copenhagen, London, Washington, and Mexico City, the meaning of the text was simply the 'mechanical' transcription of a stream of dots and dashes into an otherwise uninterpretable string of numbers and spaces or the equally mechanical conversion of a string of numbers and spaces into a sequence of dots and dashes. These clerks were employed to interpret all texts mechanically and lacked the context to interpret encrypted texts in any other way. The receipt of an encrypted text by a clerk conveyed a very large amount of information, because the clerk had few expectations about what the text might be, but the encrypted text had no intelligible meaning for the clerk who lacked the key to decrypt the message.

The staff of the German embassy in Washington obtained less information from receipt of a typescript in code 7500 than did the clerks through whose hands the message had passed, because the staff had much stronger expectations about the message. For example, the embassy staff expected the message to consist of

numbers and spaces only whereas the absence of letters was not a prior expectation of the telegraph clerks. But the message had much richer meaning for the embassy staff because they possessed code book 7500 and were therefore able to interpret the message as words in German to be re-encrypted in code 13040 using a different key. Because the staff were fluent in German they undoubtedly formed a mental interpretation (a memory) of the telegram's content that could be called upon in the subsequent investigation by German intelligence of how the message had been intercepted.

Now consider the communication of the message to Mexican President Venustiano Carranza. The German Ambassador to Mexico, Heinrich von Eckhardt, possessed a typescript (perhaps a manuscript) in German from which he made a verbal offer to Carranza in Spanish. Eckhardt's interpretation of ink marks on paper as words in German involved complex interpretation into a neural text that underwent an equally complex process of translation into Spanish in the speech production centers of his brain. The neural text from the speech production centers was then conveyed to von Eckhardt's vocal apparatus where it was composed as a phonic text (sound vibrations) that was received and interpreted by Carranza's ears and reinterpreted at various levels in Carranza's brain. At every step in the process, the constantly changing meaning was nothing more than the physical text produced by an interpreter whether this consisted of ink marks on paper, neural representations in brains (whatever thoughts may be), or vibrations in the air.

Information does not reside in things but in the reduction of uncertainty (entropy) of an interpreter who observes a thing. That which is known does not inform. If we suppose that American intelligence kept copies of German diplomatic telegrams sent from the Western Union office in Washington, then the American government learnt nothing new from the Mexican telegram provided to them by the British government beyond that the British had a copy. What provided new information was the decryption of the telegram by British intelligence into written texts in German and English. American uncertainty about whether these texts contained British disinformation was dispelled when Zimmermann conceded that he had sent the telegram and that the decryption was accurate.

The coded texts of the Zimmerman telegram were intended to be uninterpretable (meaningless) for unintended readers. This was the only reason Germany would have sent such a message via the intermediary of the U.S. State Department. The various code books were the private keys intended to be used by intended readers to interpret the strings of numbers and spaces of a typescript as an intended text in

German. British intelligence sought to reconstruct this intended interpretation without possession of the private key. Unintended information in the arrangement of code groups ('similar' words had 'similar' codes) provided clues that allowed British intelligence to construct their own key and interpret the message.

My account of the convoluted transmission of the Zimmermann telegram is reconstructed from Friedman and Mendelsohn (1994), von zur Gathen (2006), and Freeman (2006). Interpretation of what actually happened is made difficult by intentional disinformation and unintentional misinformation in the historical record. And I may have misinterpreted the texts I have read and so may be unintentionally misinforming you of the details. My words will nevertheless have served their purpose if your interpretation of what I have written approximates how I intend you to interpret my text.

## **Mutual Information and Meaning**

'Semantic information' views meaning as present in information prior to its interpretation. On this view, interpreters repackage pre-existing meaning in new media. I propose instead that meaning *be considered* the output of the interpretive process of which information is the input. On this view, answers to questions of meaning should be sought in the study of mechanisms of interpretation and the origins of interpretive competence. Once these difficult questions are answered there will be no remaining ghost in the semantics. A preference between 'semantic information' and 'meaning as interpretation' is a choice of how meaning is defined not a judgment of fact. 'Meaning as interpretation' sits comfortably with belief that information is not an objective property of things in the world but represents the epistemic uncertainty of an observer.

Consider two sequential transformations of Arthur Zimmermann's message to Heinrich von Eckhardt (Figure 2). In the first, Zimmermann's plaintext was translated into cyphertext by one of his secretaries using the textual prosthesis of codebook 7500. In the second, the cyphertext was re-encoded as an electrical signal by a telegraph clerk using the mechanical prosthesis of a transmitter. If these transformations were performed as intended, then the cyphertext could be reconstructed from the signal by a telegraph receiver and the plaintext could be reconstructed from the cyphertext by a reader in possession of codebook 7500.

Faithful transmission of the 'Zimmerman telegram' depended on *mutual information* (correlation or statistical dependency) among multiple coding systems that allowed messages to be encoded and decoded by senders and receivers with the appropriate expertise. Proponents of 'semantic information' often equate

meaning with mutual information or consider meaning to be derived from mutual information. In their view, the various texts of the ‘Zimmerman telegram’ are vehicles of a common meaning conserved through the many transformations of transmission. From the perspective of ‘meaning as interpretation,’ mutual information allows an observation of one thing to be *usefully* interpreted as *about* something else. Meaning does not reside in the immediate observation but in the synthesis of observation with background information or context in order to act effectively in the world.

There are many unrecognized statistical dependencies in the world, but these dependencies become useable only once an interpreter possesses a mechanism that ‘meaningfully’ couples observation to action. An interpreter can ‘read’ meaning from input in one of two ways. The simplest is to be given a codebook. By this means, shared possession of codebook 7500 facilitated secret communication between the Foreign Office in Berlin and the German Embassy in Washington. The second is to construct a codebook by statistical *inference* from observations in context. This was the more difficult task achieved by British intelligence and when a child learns a language.

A definition of meaning as interpretation simplifies many semantic problems. These problems include polysemy, how the same observation can have different meanings for different interpreters; synthesis, how information from multiple sources can be combined to generate new meanings; and the status of perverse but sincere interpretations. Consider divination by tea leaves. A fortune-teller observes the configuration of leaves in the bottom of a teacup and then uses the observed pattern to answer a question posed by a client. The pattern informs the fortune-teller’s advice if a different pattern would have resulted in different advice. Do all possible interpretations exist as ‘semantic information’ in the configuration of the tea leaves or does the fortune-teller mistakenly believe she has identified ‘semantic information’ that is absent? The redefinition of meaning as output avoids the horns of this dilemma and the need to distinguish ‘true’ from ‘false’ meanings. If different fortune-tellers (or the same fortune-teller at different times) interpret similar patterns as similar fortunes then mutual information exists between tea leaves and fortunes but these are meanings of the fortune-tellers not of the tea leaves.

## Information Theory and Meaning

‘The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.’ (Shannon 1948)

Shannon (1948) and Weaver (1949) developed information theory in the context of a theory of communication. Weaver (1949) conceptualized communication in ‘a very broad sense to include all of the procedures by which one mind can affect another.’ He recognized technical, semantic, and influential problems of communication: technical problems were ‘concerned with the accuracy of transference of information from sender to receiver’; semantic problems were ‘concerned with the interpretation of meaning by the receiver, as compared with the intended meaning of the sender’; influential problems were ‘concerned with the success with which the meaning conveyed to the receiver leads to the desired conduct on his part.’ Thus, Weaver used ‘information’ in the context of technical problems but ‘meaning’ in the context of semantic and influential problems. Shannon’s ‘fundamental problem of communication’ was the technical problem of reproducing a message accurately at another place. He adopted the design stance to address this technical problem and the associated mathematical theory was able to ignore problems of intentionality. Weaver, by contrast, adopted the intentional stance (‘intended meaning’, ‘desired conduct’) in talking about the semantic and influential problems, and these problems eluded mathematical treatment (see Dennett 1987 for discussion of the design and intentional stances).

In Shannon’s theory ‘the actual message [was] one *selected from a set* of possible messages.’ He used logarithmic entropy—the number of independent ways in which the message could have been different—as a measure of information content. ‘If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely.’ For Weaver, this meant that ‘The word information relates not so much to what you *do* say, as to what you *could* say. That is, information is a measure of your freedom of choice when you select a message.’ Information measured in this way had ‘nothing to do with meaning.’ For Shannon, a message possessed *meaning* if it referred to, or was correlated,



with physical or conceptual entities ‘according to some system’. This conceptualization of meaning had two components: the first was the correlation (mutual information) between messages and things in the world and the second was the ‘system’ responsible for the correlation.

Figure 3 is a variant of a famous figure of Shannon (1948) in which a message is selected by an information source and passed to a transmitter that translates the message into a signal that is transmitted to a receiver that back-translates the received signal as a message that is passed on to the destination. The signal sent and signal received differ because of inputs of unintended information (noise). The central focus of the figure is the channel between transmitter and receiver. The technical challenge was to ensure that the message passed to the destination by the receiver was as close as possible to the message passed to the transmitter by the information source. In Weaver’s example, ‘When I talk to you, my brain is the information source, yours the destination; my vocal system is the transmitter, and your ear with the eighth nerve is the receiver.’ Weaver and Shannon had relatively little to say about how information sources selected a particular message nor about how destinations interpreted the message after its receipt (nor did they have much to say about the internal workings of transmitters and receivers). In terms of my paper, information source, transmitter, receiver, and destination are all interpreters.

Weaver and Shannon recognized the *interpretation of information* to be an important question but their focus was on the *transfer of information* between interpreters. My paper addresses how information is used rather than how it is transmitted and is concerned with the general problem of interpretation (use of information) of which communication—the production and interpretation of texts (things intended to be interpreted)—is a special case. The domain of interpretation includes the use of unintended information from the environment as well as the interpretation of texts.

## **Large and Small**

‘If you make yourself really small, you can externalize virtually everything.’ (Dennett 1984, p. 143)

The study of mechanisms is frequently connected with reductionism, the ideology that larger things are properly explained by properties and interactions of their smaller parts. It is incontrovertible that events at a small scale can have large effects. A mutation in a *factor VIII* gene of a single cell resulted in at least ten of

Queen Victoria's male descendants suffering from hemophilia, including heirs to the thrones of Russia and Spain (Ingram, 1976). All genetic differences that distinguish armadillos from aardvarks and zebras from zebras originated in this way, as unintended mutations in single cells.

It should also be incontrovertible that large things affect small things. My paternal grandfather was gassed in the Great War and as he sat upright for inspection in his hospital bed, he probably raised his right hand to his head. (He performed this action when dying sixty years later in a different hospital bed). The origin of the hand salute goes back centuries, perhaps millenia, lost in the fog of time. Among many theories, the salute is the ritualized removal of a hat in the presence of a superior, the raising of the visor by a knight in armor, or a demonstration that the weapon hand is empty, but this is conjecture. John Stewart Haig's salute would have been an almost automatic gesture in the presence of an Australian or British officer but would not have been elicited by other outwardly similar persons; definitely not by a German officer. From a mechanistic perspective, his salute was caused by release of acetylcholine at neuromuscular junctions triggering actin filaments to slide past myosin filaments in muscle fibrils of his arm. But what molecular mechanism caused his hand to be raised in the presence of a person of an abstract kind? How does a military tradition move an ion across a membrane?

The detritus from collisions between protons moving in opposite directions at  $0.99999999c$  with a combined energy of 13 TeV has been interpreted as evidence for existence of the Higgs boson. Such events are created and detected at the Large Hadron Collider, housed in a circular tunnel 27 kilometers in circumference, spanning the Swiss–French border. The even more powerful Superconducting Supercollider was cancelled by the United States Congress in 1983 because of its immense cost. The differences that made the difference between the Higgs boson being detected in Europe rather than North America are to be sought in alternative political arrangements. No interpretable subatomic mechanism explains why two protons collide at these enormous energies in France but not in Texas. Events at the level of interpersonal relations and trans-Atlantic rivalries impinge upon, and predict, the movements of fundamental particles. Choices are the means by which big complex things control small simple things to make big differences.

## The Parable of the Bathtub and the Meanings of Life

Preference is an attribute of relations (differences and samenesses) rather than things (objects and events). To be told that someone chose  $x$  tells us nothing about his preferences unless we are also told what was rejected. When we choose one thing rather than another, we do not unconditionally endorse the thing chosen but express a preference for that thing relative to the other. We might be ‘making the best of a bad bargain’ or ‘choosing the lesser of two evils.’

A dichotomous choice between  $x_1$  and  $x_2$  can usefully be expressed as a choice between,  $\bar{X} + \Delta x$  and  $\bar{X} - \Delta x$  where  $\bar{X} = (x_1 + x_2)/2$  is the sameness and  $\Delta x = (x_1 - x_2)/2$  the difference. Being told the relations of things,  $\bar{X}$  and  $\Delta x$ , conveys the same information as being told their identities,  $x_1$  and  $x_2$ , but the relational form clarifies what is at stake. We choose between  $+\Delta x$  and  $-\Delta x$ , that which is different, in the context of  $\bar{X}$ , that which is the same. When complex identities are highly similar, relational forms allow efficiency of representation because the manifold samenesses need only be represented once (consider a variorum edition of a text with multiple versions).

Ponder again the flame that causes or does not cause an explosion. On one hand, the striking of the match and the presence of oxygen do not make a difference. It is the presence versus absence of hydrogen that makes the difference. On the other hand, the striking of the match and the presence of oxygen are essential parts of the mechanism that causes the explosion. When a scientist performs a controlled experiment, with and without hydrogen, she converts a possible difference into two actual series of events. (If she varied the presence of oxygen or striking of the match, then these become experimental variables and potential difference makers.) Observations are of actual things, not differences among possible things. Actual things, not differences, participate in mechanisms. But we study mechanisms to better make a difference.

Bateson (1972) defined the elementary unit of information as ‘a difference that makes a difference.’ The same words could be used to define a cause in difference-making accounts of causation (Lewis, 1973; Waters, 2007; Haig, 2014). Causation as difference-making is what could have been different. It is a history of paths not taken. Causation as mechanism is what could not have been otherwise. It is a history of a single path (one damn thing after another). The relation between these two concepts of causation is much debated

(Hall, 2004; Waters, 2007; Strevens, 2013). Different actions could have different outcomes but particular actions cause particular outcomes. One can interpret the same events as relations among identities or among differences and slip facilely between these perspectives.

An interpreter is an evolved or designed mechanism that couples entropy of observation (*possible* inputs) to entropy of action (*possible* outputs). These degrees of freedom are capabilities of the mechanism (what it *could* observe and what it *could* do). The interpreter is uncertain and undecided until an *actual* input is interpreted as *actual* action. Information is what could have been otherwise before observation. Meaning is what would have been otherwise had the observation been different. Interpretation couples information (difference maker) to meaning (difference made). For an interpreter that has evolved to intervene in its own fate, the only useful information is about differences that could make a difference.

Causation as difference-making projects our epistemic uncertainty onto the world of mechanisms, including the mechanisms of our choices. In a singular universe in which token events happen once, why should ‘could have been otherwise’ be less respectable than ‘could not have been otherwise’? What difference does it make? How can one decide? Fisher (1934) flipped a bit in my mind and I now choose ‘could have been otherwise’ but I understand neither the difference nor what is at stake. The bit *could* flip back. From the perspective of ‘could have been otherwise’, our choices *change* the world. We have evolved to choose because past choices have made a difference and future choices *will* make a difference.

An event’s causal influence can wax or wane. The ‘butterfly effect’, whereby the flapping of wings in Marilia causes a downpour in Sydney, expresses the intuition that small differences can have large effects. The existence of dynamic attractors has the opposite implication. Attractors cancel differences. Consider a bathtub. Whether water enters the tub from a spilled cup, the shower-head, the cold faucet or the hot faucet, it goes down the drain. The tub is literally a basin of attraction. All the degrees of freedom of water molecules to move ‘as they will’ come to naught because the predispositions of the tub are imposed upon them willy-nilly ‘against their will’. Each water molecule has a unique narrative of how it came to the drain but these disparate histories have no future consequences. These are differences that do not make a difference. And it is the *form* of the tub that cancels past differences, not the *matter* from which it is made. The bathtub enforces its will.

The vortex that forms as water exits a tub is a recurrent attractor. The living world abounds in recurrent goal-directed attractors because ‘endless forms most beautiful and most wonderful have been, and are being, evolved’ (Darwin, 1859). Convergent structures are attractors in adaptive space over evolutionary time. Adult forms are attractors in morphological space over developmental time. Innate and learned actions are attractors in performance space over behavioral time. Cultural conventions are attractors in social space by which a group’s members converge on common meanings. Minor variations of font or pronunciation make little difference in how words are understood. Word tokens are rocks of stability that facilitate freedom of expression. A salute is a strange attractor indeed.

Organisms embody elaborate hierarchies of homeostatic attractors that ensure basic functions are unperturbed by causal fluctuations at multiple spatial and temporal scales. Bodily attractors at all levels—from the molecular, to the cellular, to the individual—buffer organismal fates from the unforeseen. This cancellation of irrelevant differences allows ‘focus’ on what is relevant. From the myriad potential causes in its world, an organism selects those that are applicable for its needs to intervene adaptively at decision points. The organism is an ‘unmoved mover’ moved by self-selected information in pursuit of intended ends. It determines which differences will make a difference. The regress of responsibility stops here. Organisms pull their own strings.

Life is eternal recurrence. The thing that hath been shall be and that which was done shall be done. But when a sperm fuses with an egg, the nascent interpreter is something new under the sun. Why a zygote receives one set of genes rather than another can be considered chance, a concatenation of random events that picks one from the numberless possibilities that could have been sampled from its mother’s and father’s genomes. The fall of the cards cannot be anticipated, but each organism attempts to play the best hand that it can with the cards it is dealt. Each hand is new but the cards are ancient. Many designs have been tried and found wanting. Our genetic cards convey selected information from the deep past to be used with information from current events to inform our choices in a rapidly changing world. We play our hands and the cards are reshuffled.

The meaning of a life is the life that is lived. Your body is an interpretation of your life. Dualism divides this indivisible body into parts that are tools to be used once a choice has been made and parts that are texts used in choosing. When an organism acts, it is the organism’s evolutionary and developmental

history that determines to which information it responds and how it responds. The organism exerts its will as the lead actor in its own narrative, as a canceller of irrelevant causes and of competing narratives. Yet, despite this autonomy of action, many actors succumb to the slings and arrows of outrageous fortune, to factors beyond their control, foremost among which are narratives of other actors.

Self-reflective organisms (selves) respond to their world with internal changes that rewire connections between inputs and outputs to make more effective future choices; learn from experience which inputs to attend and which to ignore; perfect performance by practice with feedback from past action; and possess rich memories to inform future choice. Highly sophisticated selves augment their behaviors by observation of what works for other actors; learn from the instruction of parents and other tutors; and choose principles by which to live in pursuit of self-chosen goals. These internal changes comprise an embodied memory of the self's life-experience (the meaning of its life). This intricate and intimate private text, an interpretation intended to be self-interpreted, seamlessly melds ancient wisdom of genetic and cultural texts with news from the senses. It is responsive and responsible. It is the material and mortal soul that dies with the body.

In the beginning was mechanism. Things happened. The origin of interpretation was the origin of intentional difference-making. Choice became free as degrees of freedom of observation and intended action broadened and deepened. To understand a free choice, one must understand an interpreter's soul.

*In fine est principium.*

Acknowledgements: My text derives its meanings from many sources, poorly acknowledged. It interprets inputs from Peter Apari, Maarten Boudry, Alex Byrne, Justin Jungé, Lucas Mix, Jura Pinter, Richard Thomas, and Wenfei Tong. My debt to Daniel Dennett is obvious. Special thanks are due to a long-dead butterfly in Marilia who knew not what it did.

Borges, J. L. 1939: *The Total Library*.

Darwin, C. 1859: *On the Origin of Species*.

Dennett, D. C. 1984: *Elbow Room*. Cambridge, Massachusetts: MIT Press.

Dennett, D. C. 1987: *The Intentional Stance*. Cambridge, Massachusetts: MIT Press.

Dennett, D. C. 1991: *Consciousness Explained*. Boston: Little Brown.

Dennett, D. C. 1995: *Darwin's Dangerous Idea*. New York: Simon & Schuster.

Dennett, D. C. 2005: *Sweet Dreams*. Cambridge, Massachusetts: MIT Press.

Dewey, J. 1896: The reflex arc concept in psychology. *Psychological Review*, 3, 357–70.

Fisher, R. A. 1934: Indeterminism and natural selection. *Philosophy of Science*, 1, 99–117.

FitzGibbon, C. D. and Fanshawe, J. H. 1988: Stotting in Thomson's gazelles: an honest signal of condition. *Behavioral Ecology and Sociobiology*, 23, 69–74.

Friedman, W. F. and Mendelsohn, C. J. 1994: *The Zimmermann Telegram of January 16, 1917 and its Cryptographic Background*, revised ed. Laguna Hills, California: Aegean Park Press.

Freeman, P. 2006: The Zimmermann telegram revisited: a reconciliation of the primary sources. *Cryptologia*, 30, 98–150.

Haig, D. 2011: Sympathy with Adam Smith and reflexions on self. *Journal of Economic Behavior and Organization*, 77, 4–13.

Haig, D. 2014: Fighting the good cause: meaning, purpose, difference, and choice. *Biology and Philosophy*, 29, 675–97.

Hall, N. 2004: Two concepts of causation. In J. Collins, N. Hall and L. Paul (eds.), *Causation and Counterfactuals*. Cambridge, Massachusetts: MIT Press, 225–76.

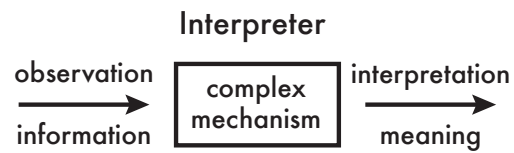
Ingram, G. I. C. 1976: The history of haemophilia. *Journal of Clinical Pathology*, 29, 469–79.

Lewis, D. 1973: Causation. *Journal of Philosophy*, 70, 556–67.

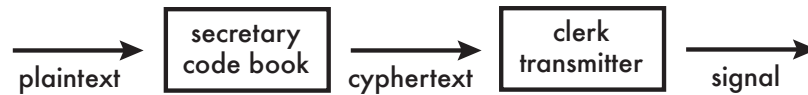
- Lindley, D. V. 2000: The philosophy of statistics. *Statistician*, 49, 293–337.
- Millikan, R. G. 1989: In defense of proper functions. *Philosophy of Science*, 56, 288–302.
- Neander, K. 1991: Functions as selected effects: the conceptual analyst’s defense. *Philosophy of Science*, 58, 168–84.
- Papineau, D. 1984: Representation and explanation. *Philosophy of Science*, 51, 550–72.
- Pearl, J. 2000: *Causality*. Cambridge University Press, Cambridge.
- Shannon, C. 1948: A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–56.
- Strevens, M. 2013: Causality reunified. *Erkenntnis*, 78, 299–320.
- Tononi, G. 2004: An information integration theory of consciousness. *BMC Neuroscience* 5, 42.
- von zur Gathen, J. 2006: Zimmermann telegram: the original draft. *Cryptologia*, 31, 2–37.
- Waters, C. K. 2007: Causes that make a difference. *Journal of Philosophy*, 104, 551–79.
- Weaver, W. 1949: The mathematics of communication. *Scientific American* 181(1), 11–15.



(Figure 1)



(Figure 2)



(Figure 3)

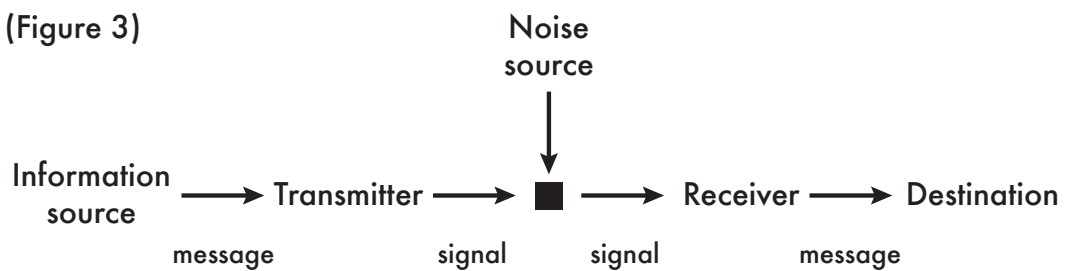


Figure 1: An interpreter is a computational mechanism for which information is input and meaning is output.

Figure 2: The interpretation of the plaintext of the Zimmermann telegram as a telegraphic signal.

Figure 3: Shannon's (1948) diagram in which a message selected by an information source is transmitted to a destination.

## Appendix: Words about words.<sup>1</sup>

‘Alle Begriffe, in denen sich ein ganzer Prozeß semiotisch zusammenfaßt, entziehen sich der Definition; definierbar ist nur das, was keine Geschichte hat.’ (Nietzsche 1887)

When I query an informant about what a thing means I frame the question as a linguistic text (spoken or written) and am answered with a linguistic text. We communicate meanings with words. For many philosophers, questions of meaning are primarily questions about language but this paper generalizes the concept of meaning to interpretations of all kinds. Linguistics and philosophy of language are vast territories of erudition into which an ill-informed novice should venture with trepidation. Nevertheless, I have been persuaded that I need to at least sketch how I would relate my account of meaning to language.

An important reason to read a text is to understand the author’s intentions. And an important reason why an author might compose a text is to have her intentions understood. Languages are elaborate conventions, shared by communities of speakers, used for the composition and interpretation of linguistic texts. Conventional meanings evolve because authors and readers often both benefit from mutual understanding of authorial intentions, but the difference between an author’s intentions and the author’s intended interpretation means that language can be used to misinform as well as inform.

This paper has defined *meaning* as the physical output of a process of interpretation and a *text* as an interpretation intended to be interpreted. By these definitions, two kinds of text are central to language that I will call public and private texts. Public texts are the strings of spoken or written words that are the outputs of language users and that are perceived and interpreted by other language users. Each language

---

<sup>1</sup> This section was included in the *Biology & Philosophy* version of the manuscript between the sections entitled *Information Theory and Meaning* and *Large and Small*. It was removed from the version submitted to *Mind & Language* to meet word limits.

user possesses a private text used in the composition and comprehension of public texts.<sup>2</sup> The private text has an intricate material form. It is a text because (1) it is a physical interpretation of the language user's life-experience in the context of innate *a priori* knowledge about language and (2) it informs the composition and comprehension of public texts. The private text is informed by the evolutionary and developmental history of the language user, especially her lived experience of public texts. Mastery of a language is the context that allows decryption of texts written or spoken in words of that language. A child learns a language in much the same way as British intelligence broke code 7500; by observation of many exemplars together with inspired guesses, tested for intelligible meaning in encounters with multiple texts, all in the context of innate and learned knowledge of how humans think.

Private texts develop over the course of a life from the cumulative perception and interpretation of public texts in the context of non-linguistic information. The forms of public texts are arbitrary conventions unique to each language but the initial bootstrapping—that allows a private text to be informed by public texts—requires key inputs from genetic texts. Public texts are designed to evoke rich associations in the private texts of intended auditors with the intention of modifying the auditors' responses. The richness of information for auditors resides in the private text not the public text. Public texts function as conventional pointers to content of private texts. This function depends on communicants sharing similar-enough private texts because of common humanity, similar life experiences, and shared membership in a linguistic community (they need to speak the same language). A speaker anticipates that a listener will interpret the public text in much the same way as the speaker would interpret the public text.

Words are defined using other words. For the writing of this paragraph, I opened a Magyar dictionary more-or-less at random and found the word *csendülni* followed by other words that I assumed to be a definition of *csendülni* in Magyar. The public text *csendülni* pointed to nothing in my private text. It evoked no associations in my mind. If you are not a Magyar speaker, there is probably little you can say about *csendülni* except that you believe it to be a word in Magyar on my dubious authority. Most Magyar speakers could probably use *csendülni* in conversation but it would have different associations for each speaker. Some might

---

<sup>2</sup> Public texts and private texts are roughly synonymous with *parole* and *langue* of Saussure (1916). Private texts are unique to each user but the needs of mutual understanding result in mutual information, and convergence, among the private texts of a linguistic community.

have no more than a vague idea what the word ‘really’ means and might consult a Magyar dictionary for the ‘correct’ definition.<sup>3</sup> (When I consult an English dictionary, I am not infrequently surprised to find that the definition differs markedly from what I thought the word meant.)

In the first chapter of *David Copperfield*, Dickens (1849) wrote: ‘In consideration of the day and hour of my birth, it was declared by the nurse ... first, that I was destined to be unlucky in life; and secondly, that I was privileged to see ghosts and spirits; both these gifts inevitably attaching, as they believed, to all unlucky infants of either gender, born towards the small hours on a Friday night.’ Gender has long been used as a synonym for sex although this sense was described as ‘now only jocular’ in the first edition of the *Oxford English Dictionary* (1899). At that time, sex was the usual term for the distinction between males and females but, within a century, sex was replaced by gender as the preferred term of many English speakers.

I have studied recent semantic shifts in the use of sex and gender as a test-case for thinking about memetic evolution (Haig, 2004). In brief, social psychologists and psychoanalysts introduced a distinction between ‘socially-constructed’ gender and ‘biologically-determined’ sex in the 1960s. Gender became a term of art in these fields, both as a way of marking a theoretical distinction and of signaling to informed listeners that the speaker believed the social to be more important than the biological. Animals had sex, only humans had gender. From this base, gender entered general discourse via its adoption by feminists of the 1980s to signal their belief in the predominant influence of social factors on sex differences, but a sex/gender distinction is now rarely maintained. Gender commonly now refers to all differences between males and females whether these are social or biological. Even hamsters have genders (Robins *et al.*, 1995). Haig (2004) suggested two factors played a role in the reconvergence of meanings of gender and sex. First, biological and social factors often interact to determine male/female differences and, in such situations, there was no neutral term. Gender became the safer default in cases of overlap, thus undermining the distinction between biological sex and cultural gender. Second, many listeners heard gender being used to refer to male/female differences but were uninformed of the theoretical distinction and thus adopted gender as the fashionable general term.

I now suspect another factor contributed to the rise of gender and concomitant decline of sex. As a student explained her preference for gender to me: ‘Gender is a category but sex is an action.’ Some

---

<sup>3</sup> My friend Apari Péter offered ‘voice of the bell’ as his spontaneous translation of *csendülni*.

speakers avoid sex because sex has copulatory connotations. Sex became a widely used euphemism for *fucking* only during the twentieth century. Did you wince when you read the ‘F-word’? I hesitated as I wrote it and still hesitate as I read it. Euphemisms are polite ways of referring to things we feel uncomfortable talking about. But, as a euphemism becomes widely adopted, our embarrassment with the subject inevitably rubs off on the euphemism which becomes less polite (consider the linguistic fate of toilet, now associated strongly with the ‘S-word’). Our hesitance in talking openly about ‘sexual activity’ has tainted sex with ‘sexual connotations’ and created a preference for a word in which these associations are less direct.

The public forms of sex and gender, as spoken and written texts, have remained unchanged through substantial changes of meaning. The important changes have been in the private associations of the public forms. Meanings of words have histories not essential attributes. At any one time, different speakers would have given different interpretations of what the public forms meant and used the public forms in different contexts. If enough speakers use a word ‘wrongly’ then their private definition becomes the accepted norm of the linguistic community. The heterodox becomes orthodox. Meaning is use. My paper attempts to persuade; to rearrange the associations of ‘meaning’ and ‘information’ in your private texts; to change how you interpret and use these words not reveal what they ‘properly’ mean.

Dickens, C. 1849: *David Copperfield*.

Haig, D. 2004: The inexorable rise of gender and the decline of sex: social change in academic titles, 1945–2001. *Archives of Sexual Behavior*, 33, 87–96.

Nietzsche, F. 1887: *Zur Genealogie der Moral*.

Robins, S. J., Fasulo, J. M., Patton, G. M., Schaefer, E. J., Smith, D. E. and Ordovas, J. M. 1995: Gender differences in the development of hyperlipemia and atherosclerosis in hybrid hamsters. *Metabolism*, 44, 1326–31.

Saussure, F. de. 1916: *Cours de Linguistique Générale*. Paris: Payot. (translated by R. Harris 1986: *Course in General Linguistics*. La Salle, Illinois: Open Court.)

Dear Professor Haig,

We have received the reports from our advisors on your manuscript BIPH-D-16-00105 "Defending the good cause: interpretation, meaning, and choice".

I am sorry to say that based on the advice received, we have decided that your manuscript cannot be accepted for publication in Biology & Philosophy.

Below, please find the comments for your perusal.

I would like to thank you very much for forwarding your manuscript to us for consideration and wish you every success in finding an alternative place of publication.

Sincerely,

Angela Potochnik  
Associate Editor  
Biology & Philosophy

Comments for the Author:

Reviewer #1: The author defends an understanding of 'meaning' as an 'interpretation' of information, which is supposed to be just the physical output of the one doing the interpreting. On this account meaning is a physical thing - an act or a page with marks, etc.

While the suggestion is an interesting one, it seems to lead to some strange conclusions. For example, if someone gets information, its meaning changes once they actually act on it. (i.e., if I get a message stating 'it is raining', the meaning is only a brain state, until I pull out my raincoat, and then the meaning is me pulling out the raincoat?) The author might think about narrowing the scope of this paper to address, say, just this understanding of information/meaning in genes, or texts, or lives and seeing if the view can be made less unintuitive in one of these arenas. If they choose to do so, it might be useful to look at and respond to other current positions and arguments in philosophy of science about information in these domains.

Also, throughout the paper, there seems to be some inconsistency in the use of the terms 'interpretation' and 'meaning'. As described above, interpretation is supposed to be a physical output, but, for example, on page 7 the author writes '...is a text intended to be interpreted as 'here is easy prey' by a predator'. How could this be an intended interpretation if an interpretation is a physical output? On page 8, 'Not only do I see the play of light on water...These meanings are my interpretations.' What is the physical output here? If is supposed to be the brain states of the observer? These seem to be different from 'hearing the clink of rigging'.

Reviewer #2: This lively paper concerns the way that meaning arises out of causation and selection in the natural world. It elaborates a view put forward by the same author in, 'Fighting the good cause: meaning, purpose, difference, and choice,' (Haig 2014, Biol Philos). Haig (2014) advanced a teleosemantic view according to which information turns into meaning when it is used by an interpreter to achieve an end, ends arising in the first instance as a result of evolution by natural selection. This new paper elaborates on what meaning is: it is the physical output of the interpreter. The author calls this 'the thing chosen' by the interpreter, while making clear that choice in this sense does not require reasoning or any intelligence on the part of the interpreter.

The paper does not make clear how the view differs from, or why it should be preferred to, existing teleosemantic accounts of meaning (e.g. Millikan 1984, Papineau 1987, Dretske 1988). The paper does go further than Haig (2014), but no clear argument is offered for the claim that meaning consists just in the physical output produced by an interpreter. Nor is it clear how misrepresentation is possible, of the kind the author subsequently discusses, on this view. The folded wings of a moth are intended to be

interpreted as \*not a moth\* by moth-predators. When a predator makes this interpretation - takes this information and uses it to inform the choice to pass on - it misrepresents (and loses out on a tasty snack). If meaning is just physical output, how can the meaning be \*not a moth\*? That is not a physical output produced by the moth predator. It is a condition that must be in place if the physical output produced by the moth predator (flying on) is to perform its evolutionary function in a normal way.

For the same reason, it wasn't clear how the frog's visual system could interpret 'incident photons on the retina as information about distance, direction of movement, and speed of the speck', if meaning consists just in the physical output produced by an interpreter; similarly for the claim that perceptions present useful information about the world (p. 6 of m/s).

An immediate problem faced by teleosemantic theories of meaning is to say why not all causal exchanges that have evolutionary functions give rise to meaning. The paper says nothing about this problem: whether the author embraces the liberal conclusion or has some way of avoiding it.

It wasn't clear what the claim that 'information is a property of relations not things' amounts to: whether it is different from the standard Shannon view, and if so, why that difference is important. Similarly, meaning is claimed to be a relation 'of things to an interpreter', not a thing in itself (p. 5 of m/s). How is an output produced by an interpreter a relation of a thing to an interpreter? Maybe the claim is: the meaning of an item of information (physical state) is a relational property of that state: it is fixed by the outputs purposefully produced by an interpreter in response to that state. However, that is not a novel claim. It is just the standard view (put forward by teleosemantic theories, amongst others).

Relations amongst things are treated exclusively as a matter of sameness and difference. No argument is given that a metaphysics of sameness and difference is adequate to capture relations of every kind, or that other kinds of relation should be excluded.

The Zimmerman telegram example suggests the opposite conclusion from the one the author draws. It strongly suggests that something is preserved across all the translations and changes of form of the telegram. They all transmit the German Empire's offer. Indeed, that is how the author presents the case (it 'conveyed an offer from the German Empire to the Republic of Mexico', p. 11 of m/s). Having presented the case, the author asserts that in the chain of transmission there was constantly changing meaning, meaning being nothing more than the physical text produced by an interpreter (p. 12 of m/s).. That conclusion does indeed follow from the author's view. But what is the argument for it? That question is especially pressing when the author's own presentation of the case suggests the opposite conclusion.

The point of the section 'Large and Small' was not clear; nor the aside on causation (p. 17 of m/s).

The author makes a nice point, 'A speaker anticipates that a listener will interpret the public text in much the same way as the speaker would interpret the public text.' This insight has the potential to avoid many of the problems generated by Gricean pragmatics and relevance semantics (Sperber & Wilson 1986). This idea has recently been developed at length by Ruth Millikan (Beyond Concepts, forthcoming).

The account of memory is suggestive and intriguing (p. 19). The internal changes which change behavioural dispositions are a kind of meaning, according to the author's account, because they are a purposeful output. Furthermore they are also intended to be interpreted, by the self, so they act as information which becomes meaningful for that interpreter (which happens to be the same self). This is an interesting kind of private meaning. Peter Godfrey-Smith has suggested that memory should be construed as the sending of messages over time (Godfrey-Smith 2014 Philos Sci). The author's view develops and complements that idea. If the author's approach to meaning were clarified and made precise, and this claim about memory formulated rigorously in terms of the author's approach to meaning, that would be a significant contribution.

The paper is replete with interesting examples from biology and other sciences, and from literature. The examples are suggestive and could be illuminating, if properly developed. Indeed, an example from this manuscript has already been quoted at length by Daniel Dennett in his latest book, *From Bacteria to Bach and Back* (2016). But without being tied to clear claims and arguments, it is hard to assess the probative value of the author's examples.

There is certainly an important role in science, and no doubt especially in philosophy, for papers that paint a vivid picture: that put forward a position clearly, without engaging in detailed arguments for it. However, the position put forward in this paper is unfortunately not explained coherently enough for it to be interpreted that way.

Dear Prof. Haig:

I write to you regarding manuscript # M&L-03-17-0180 entitled "Making sense: information interpreted as meaning" which you submitted to Mind & Language.

I am sorry to say that we are not able to accept your paper for publication in Mind & Language. At the bottom of this letter you will find some reviewer's comments. I hope that they are helpful. Please be aware that the number of submissions we receive means that we are able to publish only a small fraction of what we receive.

Thank you for considering Mind & Language for the publication of your research. I hope the outcome of this specific submission will not discourage you from the submission of future manuscripts.

Sincerely,  
David Austin  
Editorial Office

On behalf of  
The Editors of Mind & Language

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author

Unfortunately, the paper is effectively a summary with no real attempt to advance our philosophical understanding of info-theoretic conceptions of meaning/communication, or to defend such the conception against criticisms. A paper in this area revisiting the seminal work of Dretske and its influence on Fodor and many others might be worthwhile. But as the paper stands, I can see nothing to recommend its publication in a journal that has as its remit to publish novel research.

Reviewer 2

I am not sure what the point is. It is beautifully written and quite amusing, but I can't say I know what the point is supposed to be, what problem it is aiming to solve by having this conception of meaning as information. It has little to do with what I might call linguistic meaning (the pedestrian attempt to figure out how lexical terminals plus composition yield propositional meaning). It seems to be trying to argue that semantic meaning need not be "ghostly" (I didn't know that anyone thought it was just that we have no good idea of how to make it respectable). The project looks like a version of Dretske's from aeons ago, that some sort of causal account whereby coordinated physical events could "mean" by providing information. I thought that this was the standard view now (but I was a grad student a long time ago) and that the issue was not whether this COULD be so but how too make it do some non trivial work in more regular domains where meaning matters. This paper does not do this. So, I am at a loss to know what it wants to do and



whether it has succeeded in doing it. All I know is that it adds nothing to the kinds of problems I am familiar with.